A Variational Quantum Circuit Model for Knowledge Graph Embeddings

Yunpu Ma
Ludwig Maximilian University of Munich
cognitive.yunpu@gmail.com

Volker Tresp
Ludwig Maximilian University of Munich & Siemens CT
volker.tresp@siemens.com

Abstract

Can quantum computing resources facilitate representation learning? In this work, we propose the first quantum Ansatz for statistical relational learning on knowledge graphs using parametric quantum circuits. We propose a variational quantum circuit for modeling knowledge graphs by introducing quantum representations of entities. In particular, latent representations of entities are encoded as coefficients of quantum states, while predicates are characterized by parametric gates acting on the quantum states. We show that quantum representations can be trained efficiently meanwhile preserving the quantum advantages. Simulations on classical machines with different datasets show that our proposed quantum circuit Ansatz and quantum representations can achieve comparable results to the state-of-the-art classical models, e.g., RESCAL, DISTMULT. Furthermore, after optimizing the models, the complexity of inductive inference on the knowledge graphs can be reduced with respect to the number of entities.

1 Introduction

Large-scale triple-oriented knowledge databases have been proposed for knowledge representation and reasoning. These knowledge graphs (KGs) contain an increasing numbers of semantic triples and entities, with increasing computational cost in model development and inference. In this work, we propose the first quantum Ansatz for learning large-scale knowledge graphs using parametric quantum circuits and investigate its potential quantum advantages. Knowledge graphs are triple-oriented knowledge representations with semantic triples *(subject, predicate, object)* as entries. Subjects and objects are entities, represented as nodes in the graph, and predicates are labeled links. One example of a semantic triple could be *(Angela_Merkel, Chancellor_of, Germany)*. After modeling observed semantic triples of a knowledge graph, the inductive inference task is to infer the truth values of triples not contained in the training data.

2 Backgrounds

2.1 Statistical Relational Learning

We briefly introduce statistical relational learning of knowledge graphs. Let $\mathcal{E}$ denote the set of entities, and $\mathcal{P}$ the set of predicates. Let N_e be the number of entities in $\mathcal{E}$, and N_p the number of predicates in $\mathcal{P}$. Given a predicate $p \in \mathcal{P}$, the indicator function $\phi_p : \mathcal{E} \times \mathcal{E} \to \{1, 0\}$ indicates whether a triple $(\cdot, p, \cdot)$ is true or false. Let $\mathbf{a}_{e_i}, i = 1, \cdots, N_e$, be the representations of entities, and

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

$\mathbf{a}_{p_i}$, $i = 1, \cdots, N_p$, be the representations of predicates. The probabilistic model for the knowledge graphs is defined as $\Pr(\phi_p(s,o) = 1|\mathcal{A}) = \sigma(\eta_{spo})$ for all (s,p,o)-triples, where $\mathcal{A} = \{\mathbf{a}_{e_i}\}_{i=1}^{N_e} \cup \{\mathbf{a}_{p_i}\}_{i=1}^{N_p}$ denotes the collection of all embeddings; $\sigma(\cdot)$ denotes the sigmoid function; η_{spo} is the score function derived from the latent representations.

For example, in the RESCAL [1] model, entities are represented as unique R-dimensional vectors, $\mathbf{a}_{e_i} \in \mathbb{R}^R$, with $i = 1, \cdots, N_e$, and predicates are matrices, $\mathbf{a}_{p_i} \in \mathbb{R}^{R \times R}$, with $i = 1 \cdots, N_p$. The dimension R is also called the rank of the model. Moreover, the value function is defined as $\eta_{spo} = \mathbf{a}_s^\intercal \mathbf{a}_p \mathbf{a}_o$. In the TUCKER [2] tensor decomposition model, entities and predicates are real-valued vectors, with $\mathbf{a}_{e_i} \in \mathbb{R}^R$ and $\mathbf{a}_{p_j} \in \mathbb{R}^R$. By introducing a global core tensor $\mathbf{W} \in \mathbb{R}^{R \times R \times R}$, the value function in the TUCKER model reads $\eta_{spo} = \mathbf{W} \times_1 \mathbf{a}_s \times_2 \mathbf{a}_p \times_3 \mathbf{a}_o$. The computational complexity of score functions for the RESCAL and TUCKER are $\mathcal{O}(R^2)$ and $\mathcal{O}(R^3)$, respectively. Thus one goal of this work is to design a quantum Ansatz with reduced score function complexity with respect to the rank R using low-depth quantum circuits.

2.2 Variational Quantum Circuits

Here, we briefly explain the variational quantum circuit proposed in [3, 4]. A quantum circuit U with L unitary operations can be written as a product of unitary matrices $U = U_L \cdots U_l \cdots U_1$, where each unitary operation U_l could be a unitary operation acting on one qubit or a two-qubit controlled gate. In particular, a single qubit gate is a 2×2 unitary matrix in $SU(2)$, which can be parameterized as $G(\alpha, \beta, \gamma) = \begin{pmatrix} \mathrm{e}^{i\beta}\cos\alpha & \mathrm{e}^{i\gamma}\sin\alpha \\ -\mathrm{e}^{-i\gamma}\sin\alpha & \mathrm{e}^{-i\beta}\cos\alpha \end{pmatrix}$, where $\{\alpha, \beta, \gamma\}$ are tunable parameters. Similarly, the two-qubit controlled gate $C_i(G_j)$ that acts on the j-th qubit conditioned on the state of the i-th qubit can be parameterized as $C_i(G_j)\,|x\rangle_i \otimes |y\rangle_j = |x\rangle_i \otimes G_j^x\,|y\rangle_j$, where $|x\rangle_i$ and $|y\rangle_j$ are the state of the i-th and the j-th qubit, respectively.

Using the above parameterized gates, we can elaborate the quantum circuit model U_θ with parameters θ. Assume a quantum circuit with n entangled qubits, If U_l is a single qubit gate acting on the k-th qubit, then its matrix representation reads $U_l = \mathbb{1}_1 \otimes \cdots \otimes G_k \otimes \cdots \otimes \mathbb{1}_n$. Moreover, if U_l acts on the j-th qubit and conditioned on the state of the i-th qubit, then it possesses the following matrix representation $U_l = \mathbb{1}_1 \otimes \cdots \otimes \underbrace{\mathbb{P}_0}_{i\text{-th}} \otimes \cdots \otimes \underbrace{\mathbb{1}_j}_{j\text{-th}} \otimes \cdots \otimes \mathbb{1}_n + \mathbb{1}_1 \otimes \cdots \otimes \underbrace{\mathbb{P}_1}_{i\text{-th}} \otimes \cdots \otimes \underbrace{G_j}_{j\text{-th}} \otimes \cdots \otimes \mathbb{1}_n$, where $\mathbb{P}_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $\mathbb{P}_1 = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

3 Fully Parameterized Quantum Circuit Embedding for KGs

In this section we propose the *fully parameterized Quantum Circuit Embedding* (FQCE) for modeling knowledge graphs. The underlying idea is to encode the representations of entities as coefficients of quantum states, which are also called quantum representations. In this way, an R-dimensional latent representation can be encoded in an r-qubit system with $r = \lceil \log_2 R \rceil$. Quantum representations are obtained by applying parameterized quantum circuit to initial quantum states that are easy to be prepared, such that each entity is uniquely identified by the circuit architecture and the gates parameters.

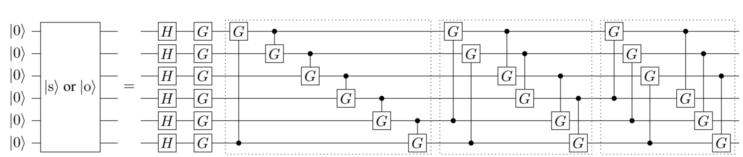

Figure 1: In FQCE, quantum representations of entities, subjects or objects, are prepared by applying unitary circuits to the initial quantum states $|00\ldots0\rangle$.

Figure 1 presents circuit architecture for generating quantum representations of all entities. For all experiments, we simulate a 6-qubit quantum system, which is initialized as a simple quantum state $|\mathbf{0}\rangle = |000000\rangle$. We then apply Hadamard gates on each qubit to create a maximal superposition state $H_6 H_5 \ldots H_1 |\mathbf{0}\rangle$. To maintain the quantum advantages, the circuit should be shallow, and the depth of the circuit needs to be in the order of $\log R$. Therefore, we only apply four additional blocks to evolve the maximally entangled state $H_6 H_5 \ldots H_1 |\mathbf{0}\rangle$, which consists of parameterized one-qubit gates and two-qubit controlled gates with control range 1, 2, and 3. The quantum representation of an entity is then encoded as the coefficients of the resulting quantum state, and it is uniquely determined by the parameters in the circuit. To be more specific, the generation of a quantum representation can be written as $U_4 U_3 U_2 U_1 H_6 H_5 \ldots H_1 |00\ldots0\rangle$. [1]

Furthermore, in FQCE each predicate p is associated with a specific quantum circuit with parameterized one- and two-qubit gates. Thus, each predicate has a unitary matrix representation $U_p(\theta_p)$, where θ_p are predicate-specific parameters from the variational quantum gates. Moreover, we fix the circuit architecture for implementing predicates as the one for preparing the quantum representations such that each predicate is uniquely determined by the circuit parameters θ_p.

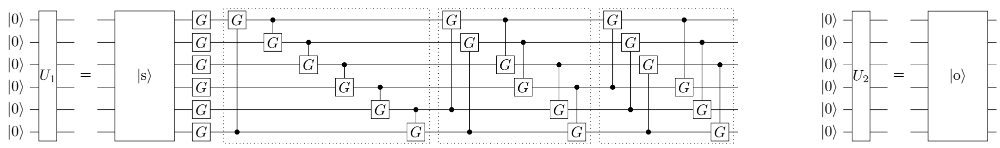

Figure 2: Building blocks of the FQCE model. Block U_1 (left) first prepares the quantum representation of the subject according to Figure 1 and evolve the state $|\mathrm{s}\rangle$ via a predicate-specific unitary circuit $U_\mathrm{p}(\theta_\mathrm{p})$. The resulting state is $|\mathrm{sp}\rangle$. Block U_2 (right) is applied to prepare the state $|\mathrm{o}\rangle$.

Given a semantic triple $(\mathrm{s}, \mathrm{p}, \mathrm{o})$, how is the score function η_spo derived in the quantum model FQCE? We define the score function as $\eta_\mathrm{spo}^\mathrm{FQCE} := \Re \langle \mathrm{o}| U_\mathrm{p}(\theta_\mathrm{p}) |\mathrm{s}\rangle$, which is the real part of the overlap of two quantum states $|\mathrm{o}\rangle$ and $|\mathrm{sp}\rangle := U_\mathrm{p}(\theta_\mathrm{p}) |\mathrm{s}\rangle$. Namely, we first create quantum states $|\mathrm{s}\rangle$ and $|\mathrm{o}\rangle$ for the subject s and object o, respectively, according to the circuit architecture given in Figure 1. We then evolve the state $|\mathrm{s}\rangle$ to $|\mathrm{sp}\rangle$ according to the predicate and evaluate the inner product between $|\mathrm{sp}\rangle$ and $|\mathrm{o}\rangle$ to obtain $\eta_\mathrm{spo}^\mathrm{FQCE}$. The block U_1 for preparing $|sp\rangle$ and block U_2 for $|\mathrm{o}\rangle$ are shown in Figure 2.

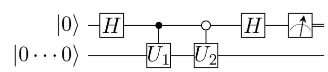

Figure 3: Quantum circuit for estimating the score function $\eta_\mathrm{spo}^\mathrm{FQCE}$. Unitary evolutions U_1 and U_2 are described in Figure 2.

In the following, we show that $\eta_\mathrm{spo}^\mathrm{FQCE}$ can be measured physically. The physical architecture is illustrated in Fig. 3, which is inspired by [5] and Observation 3 in [4]. Consider the unitary blocks U_1 and U_2, which act on the pure state $|\mathbf{0}\rangle$ conditioned on an ancilla qubit. In particular, the pure states becomes $U_1 |\mathbf{0}\rangle = |\mathrm{sp}\rangle$ if the ancilla qubit is $|1\rangle_\mathrm{A}$, and $U_2 |\mathbf{0}\rangle = |\mathrm{o}\rangle$ if the ancilla qubit is in the state $|0\rangle_\mathrm{A}$. Therefore, before applying the second Hadamard gate, the quantum state of the entire system reads $\frac{1}{\sqrt{2}} (|0\rangle_\mathrm{A} |\mathrm{o}\rangle + |1\rangle_\mathrm{A} |\mathrm{sp}\rangle)$.

Moreover, the second Hadamard gate acting on the ancilla qubit brings the system to the state $\frac{1}{2} [|0\rangle_\mathrm{A} (|\mathrm{o}\rangle + |\mathrm{sp}\rangle) + |1\rangle_\mathrm{A} (|\mathrm{o}\rangle - |\mathrm{sp}\rangle)]$. In fact, one can see that the probability of sampling the ancilla qubit in the state $|0\rangle_\mathrm{A}$ is $\Pr(|0\rangle_\mathrm{A}) = \frac{1}{2} + \frac{1}{2}\Re \langle \mathrm{o}|\mathrm{sp}\rangle = \frac{1}{2} + \frac{1}{2}\eta_\mathrm{spo}^\mathrm{FQCE}$. Hence, the score function $\eta_\mathrm{spo}^\mathrm{FQCE}$ is related to the statistics of sampled quantum states of the ancillary qubit via $\eta_\mathrm{spo}^\mathrm{FQCE} = 2\Pr(|0\rangle_\mathrm{A}) - 1$.

Similar to the classical models, this quantity defines the loss function jointly with the labels of the triplets. Given a training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^m$ with x_i being observed semantic triples, the loss function is defined as the mean error $\mathcal{L} = \frac{1}{m} \sum_{i=1}^{m} (y_i - \eta_{x_i}^\mathrm{FQCE})^{2\kappa}$, where $y_i \in \{-1, 1\}$ are labels, and $\kappa \in \mathbb{Z}^+$ is a hyperparameter. The model is optimized by updating the parameters via gradient descent. In practice, parameters of the variational gates can be estimated using a hybrid gradient descent scheme developed in [4]. In fact, according to the techniques developed in [6, 4], partial derivatives can be estimated from the statistics of the ancilla qubit using the same circuit architecture and a linear combination of gates with shifted parameters. More details of the hybrid gradient descent approach can be found in Section 4 of [4].

[1] In particular, $U_1 = G_6 G_5 G_4 G_3 G_2 G_1$, $U_2 = C_6(G_1)C_1(G_2)C_2(G_3)C_3(G_4)C_4(G_5)C_5(G_6)$, $U_3 = C_5(G_1)C_6(G_2)C_1(G_3)C_2(G_4)C_3(G_5)C_4(G_6)$, and $U_4 = C_4(G_1)C_5(G_2)C_6(G_3)C_1(G_4)C_2(G_5)C_3(G_6)$.

We briefly discuss the complexity of the model. Since we use an r-qubit system for preparing the quantum representations, with $r = \lceil \log_2 R \rceil$, and a shallow circuit with depth $\mathcal{O}(\log R)$ to represent predicates, the unitary evolution of quantum states for entities requires $\mathcal{O}(\log^2 R)$ unitary operations. The value function is estimated from the Bernoulli distribution of the ancilla qubit. Hence, one needs to perform $\mathcal{O}(\frac{1}{\epsilon^2})$ repetitions of the experiment in Fig. 3 to resolve the statistics of the ancilla qubit up to a predefined error ϵ. In summary, the evaluation of the score function η_{spo}^{FQCE} requires a run-time $\mathcal{O}(\text{poly}(\log R, \frac{1}{\epsilon}))$, realizing an acceleration with respect to the rank R.

4 Experimental Results

To evaluate the quantum Ansatz, we conduct link prediction experiments on four benchmark datasets: KINSHIP [7], FB15K-237 [8], WN18RR [9], and GDELT [10]. Negative semantic triples, not included in the datasets, are generated using the negative sampling scheme proposed in [11]. We compare the filtered ranking metrics as suggested in [11], which are filtered mean rank (MR), filtered Hits@3, and filtered Hits@10. The performance is compared with benchmark classical models re-implemented with the same embedding dimension, which are RESCAL [1], TUCKER [2], DISTMULT [12], and COMPLEX [13], and other best known results (until 2018).

Overall, the quantum circuit architecture of FQCE is fixed as in Figure 2 and 3. This 6-qubit system is simulated using a quantum circuit simulator based on Tensorflow. Since unitary evolution of a quantum state is equivalent to the unitary matrix-vector product, we can simulate and train the quantum Ansatz on a single Tesla K80 GPU without exploiting real quantum devices. Moreover, before training, all gate parameters are randomly initialized. In particular, we found that the performance of the quantum Ansatz is very sensitive to the initialization of the gate parameters. After a hyperparameter search, gate parameters are uniformly initialized in the interval $[-\pi/10, \pi/10]$.

	KINSHIP			WN18RR			FB15K-237			GDELT		
Methods	MR	@3	@10	MR	@3	@10	MR	@3	@10	MR	@3	@10
RESCAL	3.2	88.8	95.5	12036	21.3	25.0	291.3	20.7	35.1	185.0	10.4	22.2
DISTMULT	4.5	61.0	87.7	10903	21.0	24.8	305.4	23.4	39.1	130.4	12.1	24.5
TUCKER	2.9	89.8	95.0	11997	19.1	23.9	276.1	20.9	35.7	144.0	14.5	27.3
COMPLEX	**2.2**	**90.0**	**97.7**	11895	24.6	26.1	242.7	25.2	39.7	137.6	12.9	26.4
Best Known	-	-	-	4187 [9]	**44.0**	**52.0**	244.0 [9]	**35.6**	**50.1**	**102.0** [14]	**31.5**	**47.1**
FQCE	3.6	73.1	94.0	**2160**	27.4	37.8	**236.0**	19.8	33.7	131.0	10.8	24.1

Table 1: Filtered recall metrics evaluated on four different datasets.

Table 1 reports the simulation results. We can read that FQCE achieves comparable results to the classical models using the dimension $R = 64$. In some cases, e.g., the MR recall scores on WN18RR and FB15K-237, the quantum Ansatz can outperform all classical models. Another interesting observation is that the MR score on the WN18RR dataset returned by the quantum Ansatz is even better than the best-known models. However, WN18RR possesses the smallest number of average links per entities. Hence, we face the following questions. Is the quantum circuit Ansatz only practical for modeling sparse datasets with simple relational patterns due to the intrinsic linearity of the quantum circuit; and can the application of the nonlinear activation functions on quantum circuits [15, 16] further improve the performance on other dense datasets? We leave these questions for future research.

Further investigations: In the original publication [17], we also investigate different regularization methods to improve the generalization ability of the quantum Ansatz, which should simultaneously maintain the unitarity constraints required by the architecture. Besides, after visualizing the learned quantum representations via t-SNE [18], we observe a similar semantic clustering effect, namely, entities with similar semantic meaning tend to group in the vector space. Furthermore, we propose and study a quantum algorithm, which can theoretically accelerate the inference tasks on knowledge graphs, realizing a quadratic acceleration to the number of entities, namely $\mathcal{O}(\sqrt{N_e})$.

5 Conclusion

In this work, we proposed the first quantum Ansatz for statistical relational learning on knowledge graphs and introduced quantum representations. Simulations showed that the FQCE model can achieve comparable results to the benchmarks models on several datasets and realize an acceleration to the rank. Besides, we have proposed a quantum algorithm based on the FQCE building blocks, which theoretically realizes a quadratic acceleration to the number of entities for the inference tasks. One interesting **future direction** is to apply quantum representations to other NLP models.

Acknowledgement
This work has been funded by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The authors of this work take full responsibilities for its content.

References

[1] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. 2011.

[2] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[3] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

[4] Maria Schuld, Alex Bocharov, Krysta Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *arXiv preprint arXiv:1804.00633*, 2018.

[5] Pedro Chamorro-Posada and Juan Carlos Garcia-Escartin. The switch test for discriminating quantum evolutions. *arXiv preprint arXiv:1706.06564*, 2017.

[6] Andrew M Childs and Nathan Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *arXiv preprint arXiv:1202.5822*, 2012.

[7] Arthur Asuncion and David Newman. Uci machine learning repository, 2007.

[8] Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66, 2015.

[9] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. *arXiv preprint arXiv:1707.01476*, 2017.

[10] Kalev Leetaru and Philip A. Schrodt. Gdelt: Global data on events, location, and tone. *ISA Annual Convention*, 2013.

[11] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795, 2013.

[12] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.

[13] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, pages 2071–2080, 2016.

[14] Yunpu Ma, Marcel Hildebrandt, Stephan Baier, and Volker Tresp. Holistic representations for memorization and inference.

[15] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. The quest for a quantum neural network. *Quantum Information Processing*, 13(11):2567–2586, 2014.

[16] Eric Torrontegui and Juan José Garcia-Ripoll. Universal quantum perceptron as efficient unitary approximators. *arXiv preprint arXiv:1801.00934*, 2018.

[17] Yunpu Ma, Volker Tresp, Liming Zhao, and Yuyi Wang. Variational quantum circuit model for knowledge graph embedding. *Advanced Quantum Technologies*, 2(7-8):1800078, 2019.

[18] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.